REPÚBLICA ARGENTINA
UNIVERSIDAD DEL SALVADOR
FACULTAD DE CIENCIAS ECONÓMICAS
MAESTRÍA EN AUDITORIA DE SISTEMAS

Aplicar Data Mining en la Detección y Prevención de Fraude con Tarjetas de Crédito y Débito

TESIS

PARA OPTAR AL GRADO DE:
MAGÍSTER EN AUDITORIA DE SISTEMAS

Autor : Ing. Nelly Tintaya Maquera

Tutor : MBA. Daniel Nocella

Buenos Aires, Marzo de 2010

A Dios, quien día a día, me ha dado fuerza para llevar a cabo todo lo que me he propuesto en mi vida.

A mis Padres Apolinar y Felipa quienes siempre me han guiado y apoyándome siempre en los éxitos o fracasos, quienes me han enseñado a diferenciar lo bueno de lo malo, los que siempre están pendiente de mí cuando estoy cerca y lejos, a ustedes, los quiero con toda mi alma

AGRADECIMIENTOS

A mi Papá, Mamá y a mis hermanos, por todo el amor y fortaleza que me han dado, porque cada uno de ustedes forma parte de lo que soy, esta tesis es una superación no solo personal sino también familiar, sin ustedes jamás lo hubiera logrado.

A mi hermana Eugenia por todos los juegos, conversaciones y momentos vividos, aún en la distancia, a mis hermanos Apolinar y Juan Carlos quienes mantuvieron una constante preocupación por mis estudios y yo agradezco su interés por mi futuro.

Muy especialmente a los funcionarios de las empresas e instituciones que colaboraron respondiendo la encuesta, pues sus apreciaciones contribuyeron al mejoramiento de la propuesta que aquí se presenta.

A mi asesor de Tesis, el MBA. Daniel Nocella, por todos sus conocimientos y experiencia brindados para alcanzar los objetivos de trabajo.

A las diferentes personas (familiares, compañeros de trabajo y amigos en general) quienes con su pregunta ¿cómo va la tesis?, siempre me dieron el impulso y me comprometieron para seguir adelante hasta alcanzar esta meta.

RESUMEN

El objetivo ultimo de la tesis consiste en demostrar que mediante Data Mining puede diferenciarse claramente el comportamiento de clientes normales de tarjetas crédito. En este trabajo, el análisis se realizo mediante métodos de clustering y Redes Neuronales de una serie de datos sobre el titular de la tarjeta y la transacción que ha permitido.

El desarrollo de la tesis, en tanto, consiste en detallar paso a paso con una metodología de data mining (CRISP-DM) como formular un plan para enfocar las campañas bancarias mediante un segmentación de riesgo y rentabilidad basada en modelos predictivos generados a partir de los datos de las bases operacionales antes mensionadas. Finalemente, la principal conclusión que se obtiene como resultado de la investigación consiste en que la tesis planteada en un comienzo es valida, por cuanto es posible construir un modelo como el deseado.

La investigación desarrollada no ha pretendido elaborar modelos predictivos específicos, puesto que para ello hubiese sido necesario disponer de datos concretos de un conjunto masivo de clientes, información cuyo acceso esta severamente restringido por la ley de secreto bancario. Por lo tanto, las indicaciones que se entregan en este informe son genéricas, puramente conceptuales, no están dirigidas a ningún Banco especifico y no se basan en ningún conjunto de datos de ningún grupos de personas.

Se ha empleado la metodología CRISP-DM para la elaboración del plan de data mining por corresponder a un estandar ampliamente utilizado en proyectos de data mining. Por otro lado, dado que hubiese sido imposible ilustrar el plan de data mining sin recurrir a alguna herramienta de data mining especifica, se ha usado WEKA por hacer posible la representación de flujos de datos de manera grafica y por su gran acidad de trabajar con una amplia gama de bases de datos operacionales de distintos proveedores.

ÍNDICE DE CONTENIDO

CAPITULO 1. INTRODUCCIÓN

CAPITULO 2. MARCO TEORICO

INDICE DE FIGURAS

INDICE DE TABLAS

CAPÍTULO 1

INTRODUCCIÓN

En este capítulo se describe el contexto de la tesis, la problemática del entendimiento del negocio seguido de los objetivos de la misma y finaliza con su estructura.

1.1 ALGUNAS CONSIDERACIONES

El fraude Bancario evoluciona día con día, los defraudadores utilizan métodos más sofisticados y tecnología de punta para realizar transacciones fraudulentas. Esta tendencia ha obligado a las instituciones financieras alrededor del mundo a desarrollar estrategias de valor para disminuir el fraude y mitigar el impacto que representa para el negocio una práctica cada vez más compleja y generalizada alrededor del mundo, ya que diariamente se genera millones de transacciones electrónicas en el sistema financiero usando tarjetas de crédito y debito, donde un porcentaje de ellas son transacciones fraudulentas.

El uso fraudulento de tarjetas de crédito supone un coste de miles de millones de dólares anuales para el sistema bancario y la economía mundial. Pese a las numerosas medidas ensayadas para combatirlo, la cantidad y sofisticación de este tipo de delitos aumenta cada año, superándose sistemáticamente las medidas anti-fraude.

En la actualidad, por lo general las entidades financieras no poseen herramientas eficaces, en la mayoría cuentan con sistemas de detección de fraude a posteriori, muy pocas cuentan con sistemas de detección de fraude a priori, basados en data minig para detectar transacciones fraudulentas.

Este estudio analiza la eficacia de las técnicas procedentes de la computación flexible para apoyar la toma de decisiones estratégicas, generando una metodología orientada a mantener vigente el conocimiento del negocio basada en la extracción de patrones desde las bases de datos transaccionales que la misma empresa genera a diario pero no utiliza con este propósito por no contar con metodologías eficientes.

La competencia globalizada, los ciclos de negocios son cada vez mas breves, esto exige agilidad y rapidez para que las decisiones sean oportunas, todo ello aumenta las condiciones de imprecisión y la incertidumbre ya que las fuentes de información para las variables que intervienen en las decisiones estratégicas presentan diferentes dificultades, entre otras, las causadas por la complejidad y el gran volumen de transacciones almacenadas en las bases de datos y por la imperfección que estos datos presentan para los fines de gestión.

El ámbito de estudio corresponde a entidades emisoras de Tarjeta de crédito, para ello se ha trabajado con las grandes tiendas o Puntos de Ventas. En el contexto de este trabajo , surge la oportunidad de desarrollar la metodología propuesta, que permite superar las dificultades del actual escenario del área Gestión y Riesgo, utilizando los grandes avances tecnológicos en la informática, en particular en la extracción de conocimiento desde grandes bases de datos.

Generalmente, entidades emisores disponen de sistemas que realizan algún tipo de comprobación de las transacciones, utilizando sencillas reglas *condicionales*. El problema de estos sistemas es que, aunque intuitivamente se sepa que ciertas reglas detectan el uso irregular de una tarjeta, normalmente resulta imposible expresarlas con validez empírica. En consecuencia, el banco a menudo se enfrenta al dilema de identificar erróneamente una tarjeta como fraudulenta cuando en realidad no es el caso, lo que implica el riesgo potencial de deteriorar la relación con el cliente.

Aplicando técnicas de data mining, puede diferenciarse claramente su comportamiento del de los clientes normales. Definir las características que, combinadas, caracterizan los diversos tipos de fraude. Diferenciar el uso fraudulento del normal, entre los diversos modelos y técnicas existentes en Data Mining, se ha elegido las Redes Neuronales, las cuales, nos permiten detectar y analizar las múltiples relaciones entre los muchos elementos de datos simultáneamente para detectar patrones que formar un panorama completo de la tarjeta - titular de la conducta-y al instante identificar inusual comportamiento

1.2 PLANTEAMIENTO DEL PROBLEMA

Actualmente algunas entidades financieras conocen el fraude después de su realización por medio de reclamos del cliente o porque había un sistema de detección de fraude a posteriori, es decir, se detecta el fraude pero lastimosamente ya no se puede recuperar el monto de dinero perdido producto de la transacción fraudulenta

En este trabajo se aborda el problema de la detección de cambios de consumo de usuarios de tarjetas de crédito fuera de lo normal, y la correspondiente construcción de estructuras de datos que representen el comportamiento reciente e histórico de cada uno de los usuarios, teniendo en cuenta la información que contiene una